AI predicts grassland SOC, but still falls short of explaining its dynamics

Jianxiao Su,[1,2] Yishu Yang,[1,2] Mengyao Yu,[1,2] and Jie Gao[1,*]
[1]Key Laboratory for the Conservation and Regulation Biology of Species in Special Environments, College of Life Science, Xinjiang Normal University, Urumqi 830054, China
[2]These authors contributed equally
*Correspondence: jiegao72@gmail.com
Received: April 22, 2026; Accepted: July 14, 2026; Published Online: July 16, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100058

Citation: Su J., Yang Y., Yu M., et al. (2026). AI predicts grassland SOC, but falls short of explaining its dynamics. *The Innovation Informatics* **2**:100058.

Grassland soil organic carbon (SOC) stocks are central to terrestrial carbon sequestration and climate mitigation.[1] Artificial intelligence has improved the ability to generate spatially continuous estimates of grassland SOC stocks across landscapes and regions. By integrating field observations with remote sensing products, environmental covariates, and machine-learning algorithms, recent studies have strengthened broad-extent SOC stock mapping and carbon-cycle quantification across heterogeneous grassland systems.[2,3] This is a substantial advance for grassland research, where field observations are often sparse and environmental heterogeneity is high. To avoid conflating these terms, we distinguish SOC prediction from SOC mapping and mechanistic explanation. Here, prediction refers to the statistical estimation of SOC stocks from observed covariates, whereas mapping refers to the

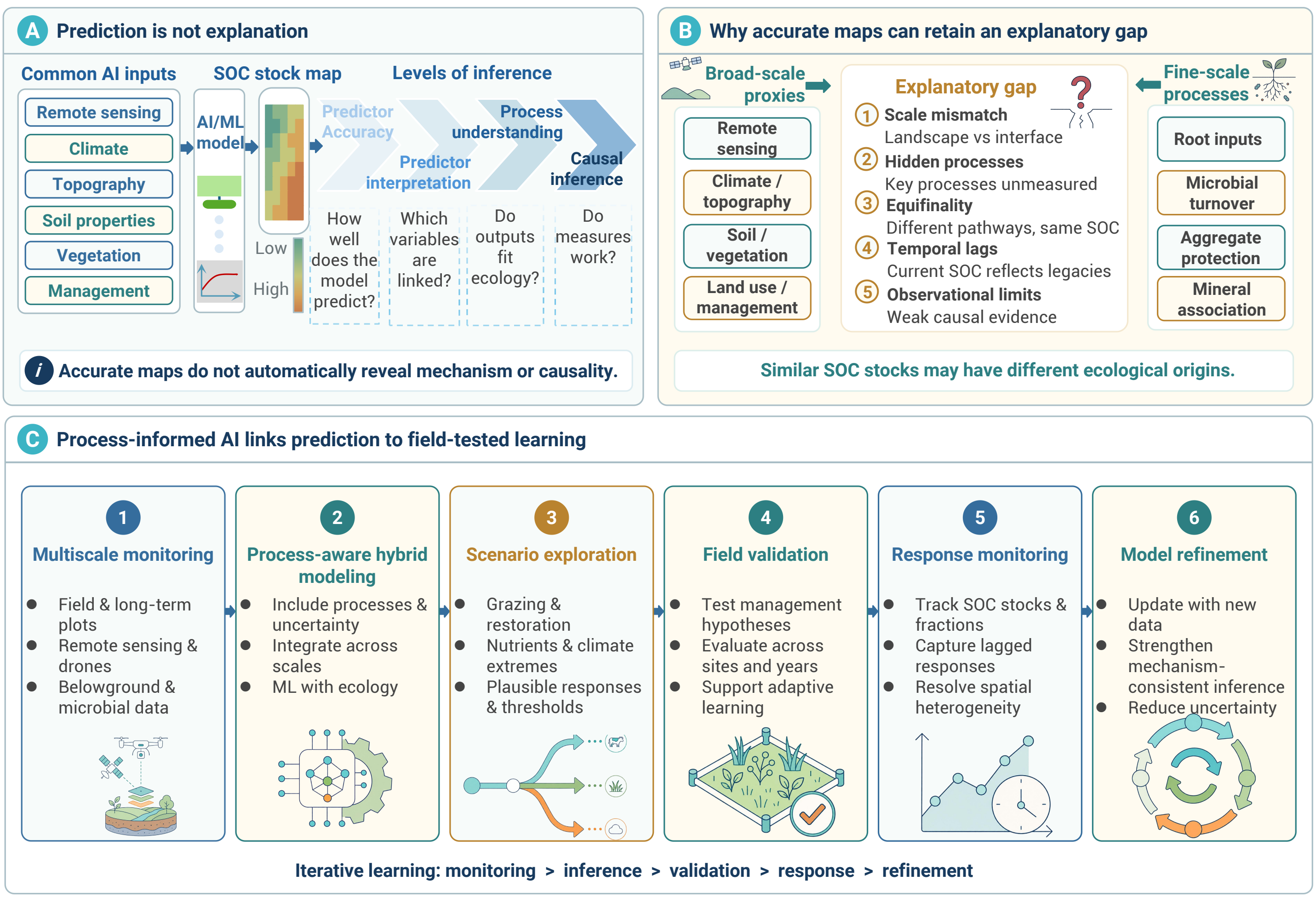

Figure 1. An inferential framework for AI-based grassland SOC research The upper panel distinguishes prediction, interpretability, mechanistic consistency, and causal explanation. The middle panel summarizes major sources of the explanatory gap, including scale mismatch, hidden variables, equifinality, temporal lags, observational training data, and limited field validation. The lower panel outlines a process-informed learning loop linking monitoring, hybrid modeling, scenario evaluation, field validation, and model updating.

production of spatially continuous SOC stock estimates. Explanation is used more narrowly to mean inference about the processes that form, transform, stabilize, or deplete SOC, including carbon inputs, microbial transformation, aggregation, and mineral association. Yet more accurate SOC stock maps should not be taken as equivalent to stronger mechanistic or causal explanations. Prediction, interpretability, mechanistic consistency, and causal explanation represent related but distinct levels of inference. Accurate SOC prediction can reveal robust spatial covariation; interpretable machine-learning tools can identify influential predictors and interactions; and process-informed modeling can evaluate whether these patterns are consistent with known pathways of carbon input, microbial transformation, and stabilization. Causal explanation, however, requires stronger evidence, including process measurements, temporal observations, experimental manipulation, or model structures that explicitly represent causal pathways. The challenge is therefore not to replace prediction with explanation, but to move grassland SOC modeling along this inferential gradient. The upper panel of Figure 1 summarizes this inferential gradient by distinguishing predictive accuracy, model interpretability, mechanistic consistency, and causal explanation. Many current AI frameworks can identify where SOC stocks are high or low, but they remain less effective at resolving how SOC is formed, why it persists, and under which conditions it is lost.

The reason is straightforward. Most AI workflows treat SOC stocks as response variables and relate them to a set of environmental covariates. This is entirely sensible when the objective is predictive mapping of SOC stocks. It is less sufficient when the objective is to infer the ecological processes that generate those stocks, and still less sufficient when the goal is causal explanation. SOC is not a single process or even a direct state variable in any simple sense. It is the cumulative outcome of carbon inputs, microbial processing, and stabilization.[1,4] More roots, more litter, more rhizodeposition, slower decomposition, stronger aggregation, or tighter mineral association can all lead to a higher SOC stock, but they do not mean the same thing. They imply different controls, different vulnerabilities, and different management options. When all of these pathways are compressed into a stock map, the model may perform well while the mechanism remains unresolved.

The problem is especially acute in grasslands because a large share of the carbon pathway is belowground. Root turnover, exudation, microbial growth efficiency, aggregate formation, and mineral protection all contribute to whether carbon is retained or released.[4] Similar SOC stocks may therefore arise from very different ecological histories. One site may accumulate carbon because belowground inputs are sustained; another because decomposition is constrained; and a third because carbon is more effectively protected in aggregate or mineral-associated pools. SOC stock estimates alone cannot distinguish among these pathways.

The middle panel of Figure 1 identifies several sources of the explanatory gap between accurate SOC stock mapping and stronger mechanistic or causal inference. One source is the mismatch between broad-extent AI-ready predictors and the fine-scale processes that regulate SOC formation and persistence. AI models often perform well with landscape- to regional-scale observations, and remote-sensing products provide broad coverage, repeated observations, and consistent predictors.[3,5] By contrast, many mechanisms that govern SOC dynamics operate at the rhizosphere, aggregate, and mineral-interface scales. Climate, vegetation cover, topography, and land use therefore often serve as proxies for direct observations of root-derived inputs, microbial turnover, aggregate protection, or mineral association.

This gap, however, is not only spatial. It also reflects hidden or incompletely measured variables, equifinality, temporal mismatch and lags, the observational nature of most training data, and limited field validation of management interventions. Similar SOC stocks may arise from different combinations of carbon input, microbial processing, decomposition constraints, aggregation, and mineral protection. Current SOC stocks may also carry legacy effects of past climate, vegetation, grazing, degradation, or restoration. Because most models are trained on observational covariates rather than on process-resolving measurements, time-series observations, manipulative experiments, or intervention-validation data, high predictive accuracy does not, by itself, identify the causal pathways that generated the observed SOC pattern.

For that reason, the next step is not simply more prediction nor a rejection of predictive AI. A more useful direction is to link predictive modeling more explicitly with process observation, ecological theory, and field validation. Remote sensing and drones remain essential, but they need to be more tightly linked to field sampling, long-term plots, belowground observations, and microbial measurements. If the aim is to understand SOC dynamics rather than only produce SOC stock maps, models need information on the processes they are expected to explain. Root inputs, microbial biomass and turnover, particulate and mineral-associated carbon fractions, and aggregate structure are therefore part of the system itself, not optional additions.[1,4]

Process-informed AI also requires a change in modeling logic. The question should not only be which variables best predict SOC stocks, but also what those variables represent. Some variables may indicate carbon inputs, others microbial transformation, stabilization, or management pressure. Hybrid approaches are useful because machine learning can capture nonlinear responses and interactions, whereas ecological knowledge helps structure interpretation.[2] Together, they provide a more realistic route toward stronger mechanistic inference.

Building on this logic, the lower panel of Figure 1 frames process-informed AI less as an immediate decision engine than as a tool for adaptive learning and field validation. Once monitoring is linked to process-aware and hybrid modeling, AI can be used to generate management hypotheses and evaluate plausible scenarios, such as grazing intensity and timing, restoration practices, nutrient addition, or responses to climate extremes.[5] These outputs should be treated as decision support rather than direct prescription. Grassland SOC responses are often slow, lagged, and spatially heterogeneous, so scenario outputs need to be tested through field validation, repeated monitoring, and model updating before they can provide robust guidance for management. In this sense, process-informed AI can strengthen mechanistic inference and improve scenario evaluation, but it does not remove the need for causal evidence from process observations, temporal data, manipulative experiments, and field-based intervention tests.

The promise of AI in grassland SOC research is therefore real, but it needs to be stated more carefully than it often is. AI has already improved our ability to estimate SOC and compare its spatial distribution across grasslands. That is not trivial. But if SOC continues to be treated mainly as a static stock to be fitted, then even very accurate models will have limited explanatory value. The real opportunity is to use AI to connect SOC maps with process observations, ecological theory, and evidence-based decision support. When that happens, AI will not simply make better maps of grassland SOC. It will help identify why those patterns emerge and under what evidence-supported conditions they might be changed.[1-5]

REFERENCES

1. Bai Y. and Cotrufo M.F. (2022). Grassland soil carbon sequestration: Current understanding, challenges, and solutions. *Science* **377**:603–608. DOI:10.1126/science.abo2380
2. Demattê J.A.M., Poppiel R.R., Novais J.J.M., et al. (2025). Frontiers in earth observation for global soil properties assessment linked to environmental and socio-economic factors. *The Innovation* **6**:100985. DOI:10.1016/j.xinn.2025.100985
3. Zhao T., Wang S., Ouyang C., et al. (2024). Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation* **5**:100691. DOI:10.1016/j.xinn.2024.100691
4. Tao F., Huang Y., Hungate B.A., et al. (2023). Microbial carbon use efficiency promotes global soil carbon storage. *Nature* **618**:981–985. DOI:10.1038/s41586-023-06042-3
5. Miao Y., Tong R., Zhu N., et al. (2025). Effects of stand structural diversity on carbon storage of Masson pine forests in Fengyang Mountain Nature Reserve, China. *For. Res.* **5**:e011. DOI:10.48130/forres-0025-0010

FUNDING AND ACKNOWLEDGMENTS

We sincerely thank the editors and reviewers for their time and careful evaluation of this manuscript. This work was supported by the National Natural Science Foundation of China (grant no. 32401662) and the Young Outstanding Talents Project of Xinjiang Normal University (grant no. XJNUQB2023-14). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Jianxiao Su wrote the initial draft, while Yishu Yang and Mengyao Yu provided diagramming ideas and suggestions. Jie Gao offered important revisions and academic guidance. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.